PERGAMON

Computers & Geosciences 29 (2003) 655–664

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

MDA: a MATLAB-based program for morphospace-disparity analysis[☆]

Nicolas Navarro*

UMR CNRS 5561, Biogéosciences, Centre des Sciences de la Terre, Université de Bourgogne, 6 bvd Gabriel, Dijon F-21000, France

Received 14 June 2002; received in revised form 3 December 2002; accepted 15 December 2002

Abstract

A MATLAB® program that examines patterns of state-space occupation is described. Four subroutines are available with which to visualize morphospace patterns: (i) in terms of their features such as dispersion, aggregation and location, thereby allowing users to extract complementary quantitative information about how the state-space is structured, and (ii) in terms of changes in those patterns that can be compared with other biotic (e.g., extinction, origination rates) or abiotic (e.g., environmental proxy) information. The program incorporates many of the latest and most widely used statistical parameters for describing multivariate spaces. The parameters are estimated on the basis of bootstrap resampling or bootstrap rarefaction procedures. Applications based on stochastic simulation of the evolution of monophyletic clade (using m-file contained in the help folder of the MDA program) are presented so as to illustrate the program's various options. The versatility of MDA allows the most interesting patterns to be extracted rapidly from data and the program to be applied readily to a wide range of state-space problems.

Keywords: MATLAB; State-space; Disparity; Paleobiology; Bootstrap

1. Introduction

Over the last decade, many macroevolutionary studies have focused on the development and the utilization of the morphospace-disparity framework (Foote, 1991, 1993, 1994, 1999; Wills et al., 1994). Likewise, the utilization of morphospace in microevolutionary studies has become widespread (de Vargas et al., 2001; Norris, 2000; Renaud et al., 1996, 1999). The notion of morphospace is related to (i) representation of elements of whole organism, or (ii) representation of samples of overall morphology in a uni- or multivariate state-space (Eble, 2002b). However, considering the nature of data, several dichotomies have been underlined and sometimes contrasted and various concepts have emerged such as theoretical and empirical morphospaces (McGhee, 1999; Eble, 2000b, 2002a), hybrid morphospace (McGhee, 1999), developmental morphospace (Eble, 2002a), or theoretical design space (Hickman, 1993).

The methodological framework of morphospace-disparity has been used to track changing patterns of morphological differentiation among taxa and to evaluate the influence of developmental and morphological constraints on evolutionary radiations (Eble, 2000a, 2002a; Foote, 1991, 1993, 1994, 1997, 1999; Jernvall et al., 1996; Lupia, 1999; Smith and Lieberman, 1999; Wills et al., 1994). This framework seems readily transposable to the identification of patterns of morphological change in species lineages. In both cases, these analyses provide a time signal that can then be compared with other intrinsic (e.g., diversity) or extrinsic patterns (e.g., environmental proxy) and so

---

[☆] Code available from http://www.u-bourgogne.fr/BIOGEOSCIENCE/NNavarro/MDA_prgm.html, or http://www.iamg.org/CGEditor/index.htm

*Tel.: +33-3-8039-6382; fax: +33-3-8039-6387.

E-mail address: Nicolas.Navarro@u-bourgogne.fr (N. Navarro).

0098-3004/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0098-3004(03)00043-8

guide the search for potential causal factors. The morphospace-disparity framework is currently a major area of inquiry in paleobiological studies.

All of the routines used in developing the morphospace-disparity concept have been user-programmed and intended for specific purposes (i.e., just a few descriptors used for just one type of analysis). This MDA program, however, proposes four major modes of investigation incorporating the majority of estimators of morphospace occupation (location and dispersion) found in the literature. MDA runs under MATLAB® with a simple text user-interface. MATLAB® is a product of MathWorks, Inc. that integrates a high-level language based on functions, mathematical computing, and visualization providing a flexible environment with an open architecture for data exploration.

Although it was developed initially for investigating morphospaces in macroevolutionary theory, the MDA program can be extended to other areas of research in the geosciences or biosciences that use the concept of state-space for investigating data, such as ecological spaces (Jernvall and Wright, 1998), environmental spaces (Jackson and Overpeck, 2000), geochemical spaces and others.

2. Disparity metrics, estimation and analysis

2.1. Disparity metrics

MDA routines are designed to operate either on continuous variables obtained from multivariate ordination—such as principal components analysis, principal coordinates analysis (PCO), and correspondence analysis—or with the raw morphospace (i.e., without ordination) of continuous characters. Procedures for studying discrete characters such as character-state combinations (Foote, 1995, 1999) are specific and problematic to compute. Therefore, no procedure has been included for studying such characters. However, a set of discrete characters can be analyzed by using scores obtained previously from a PCO of a Gower-transformed distance matrix (e.g., Foote, 1994, 1999).

For continuous data, numerous methods of quantifying morphospace occupation have been proposed: mean Euclidean pairwise distance (Wills et al., 1994), median pairwise distance (Wagner, 1997) or the mean Euclidean distance between specimens and the subspace centroid (Wills et al., 1994), the sum of the square distances between specimens and the overall centroid divided by $(N-1)$ (Foote, 1993), the morphospace occupation counted as cells occupied in subspace (Foote, 1992; Dommergues et al., 1996), the sums of variances or ranges (Foote, 1991; Wills et al., 1994; Eble, 2000a), range as the maximum Euclidean pairwise distance (Ciampaglio et al., 2001), hyper-cuboid or hyper-ellipsoid volume (Wills et al., 1994; Foote, 1999), convex hull volume (Foote, 1999), and PCO volume (Ciampaglio et al., 2001). Many of these metrics are available with MDA (Table 1). However, despite this wealthy literature about methodological development of metrics and their applications, few have really systematically compared the behavior of the different metrics to variation in the data structure (see Ciampaglio et al., 2001).

Alternatively, various distributional parameters (mean, median, modes, skews, minimum and maximum) have been used to detect and distinguish long-term trends in the evolutionary history of clades (McShea, 1994; Roy et al., 2000; Wagner, 1996). Minima and maxima quantify the extreme poles of morphospace occupation, while other parameters reveal events relating to the central trend. Accordingly, minima and maxima reveal the expansion or contraction of space occupation. So, in a macroevolutionary framework, say, they correspond to morphological innovations resulting from the appearance of new species, or to the loss of morphotypes resulting from species extinction. Hence, in addition to the 10 metrics relating to dispersion, two additional estimates of location are available: they are the minimum and the maximum of each component. For examples of how to utilize these parameters see McShea (1994).

2.2. Bootstrap, rarefaction and estimators of metrics

Because observed values of disparity metrics are sample dependent and are biased by the experimental procedure, an estimator and its associated standard error may be computed by using bootstrapping (Efron and Tibshirani, 1993): data are randomly resampled with replacement, and the mean and standard deviation are then calculated. The standard deviation then provides an estimate of standard error, which, in paleobiological studies, essentially reflects an estimate of analytical error (Foote, 1993; Eble, 2000a). The user chooses whether or not to calculate a confidence interval (CI) for estimators. This CI is based on the percentile method. The user determines the number of bootstrap resampling iterations (no default number of replicates is predefined). When using the mean with its standard deviation as an estimator of standard error, 200 replicates may be a "good" minimum value (standard deviation tends to have stabilized at this value, but this depends on the data), and this number may be increased if a CI is chosen. For the bootstrap tail test (see below), at least 1000 replicates are required to achieve meaningful accuracy.

Disparity estimates can be grouped into two categories: range-based estimates and variance-based estimates (Table 1). Estimates from the first category and the two location parameters (minimum and maximum

Table 1
Disparity estimates available in MDA version 1.2

Disparity estimates		Analysis	Rarefaction	Truncation limit[a]	References
Range-based					
Sum of ranges	Sum of univariate ranges	SGA, MGA	*y*	*n*	Foote, 1991
Root-product of ranges	*N* root of product of univariate ranges	SGA, MGA	*y*	*n*	Foote, 1991
Range	Range considered as the maximal Euclidean pairwise distance	SGA, MGA	*y*	*n*	Ciampaglio et al., 2001
Area of convex hull	Area of the polygon described by vertex of the convex hull	SGA, MGA	*y*	2	Foote, 1999
PCO Volume	of the number of species	SGA, MGA	[b]	*n*	Ciampaglio et al., 2001
Variance-based					
Sum of variances	Sum of univariate variances	SGA, MGA, BTailTest	*n*	*n*	Foote, 1991
Root-product of variances	N root of product of univariate variances	SGA, MGA BTailTest	*n*	*n*	Foote, 1991
Mean pairwise distance	Mean of the Euclidean pairwise distance	SGA, MGA, BTailTest	*n*	*n*	Wills et al., 1994
Median pairwise distance	Median of the Euclidean pairwise distance	SGA, MGA, BTailTest	*n*	*n*	Wagner, 1997
Mean distance to centroid	Mean of the Euclidean distance between species and centroid of the subspace	SGA, MGA, BTailTest	*n*	*n*	Wills et al., 1994
Partial disparity	species and overall centroid divided by (N-1)	PDA, BTailTest	*n*	*n*	Foote, 1993
Location					
Minimum and maximum	Minimum and maximum of each component	SGA, MGA	*y*	10	McShea, 1994

[a] Some estimates are limited either in their computation (convex hull) or in results (Minimum and Maximum).

[b] Rarefaction is performed by standardization by sample size (see text for discussion).

on each component) are sensitive to sample size (Foote, 1992). To correct sample size dependence, an optional rarefaction procedure is used (see Foote, 1992; Wills et al., 1994; Eble, 2000a). For the majority of estimators, rarefaction is carried out by bootstrapping. The bootstrapping procedure used for rarefaction yields the same statistical structure of error for all indices, whether they are rarefied or not (Eble, 2000a). In contrast, for PCO volume (Ciampaglio et al., 2001), rarefaction is performed using standardization by sample size following the initial procedure of Ciampaglio et al. (2001). Results obtained are similar to those obtained using rarefaction by bootstrapping. The code can be easily modified to use bootstrapping rarefaction instead of sample size standardization for the PCO volume. For the rarefaction procedure, the user can select the sample size or, alternatively, can use a predefined sample size that corresponds to the minimum size observed.

Another procedure is incorporated in the first steps of MDA that allows axes to be rescaled to their eigenvalues. This option may be required where, say, variance for each component has been scaled to unity (dividing the initial scores by the root squared of the eigenvalue) as implemented in principal component analysis (PCA) by certain statistical packages. By rescaling with the root squared of eigenvalues, their initial proportion of variance can be recovered. Such rescaling is unnecessary when data are derived from PCA with other programs (such as many PCA function in MATLAB), PCO, correspondence analysis or from partial warps, because variances on the axes are then equal to their respective initial proportion of variance.

2.3. *Analyses performed*

Four modes of investigation are available: analysis of the temporal pattern of morphological occupation for a single—SGA—or multiple groups—MGA—, decomposition of disparity between groups, i.e., partial disparity (Foote, 1993)—PDA—and a bootstrap tail test—BTailTest—(Efron and Tibshirani, 1993).

SGA and MGA are the most common studies (e.g., Foote, 1991, 1999; Eble, 2000a). Morphospace is subdivided by temporal level (and/or geographical, taxonomic group or other criteria) and the disparity of each subspace is estimated. MGA runs along the same lines as SGA but covers more than one group.

PDA analysis corresponds to a decomposition of variance (Foote, 1993). As with MGA, morphospaces are subdivided by temporal level and taxonomic groups. In this case, a specific metric is used (Table 1). For each temporal subspace, the overall disparity is divided amongst the different groups existing at that time (e.g., Foote, 1993; Smith and Bunje, 1999).

The last analysis option corresponds to a bootstrap tail test. This test has been used for disparity studies in various situations such as the effect of clade omission on disparity of Trilobita (Foote, 1993), or the origin of orders of atelostomate echinoids (Eble, 2000a). Observed disparity estimates of the groups under study are compared with the bootstrap distribution of the disparity estimate of the reference group. Because the number of replications must be high enough to ensure accurate results, this option is limited to one disparity estimate. This estimate is chosen by the user in the available variance-based metrics (Table 1). Range-based metrics are unavailable because their unbiased estimation (without sample size dependence) requires rarefaction. Thus, observed value of a range-based metric is biased by sample size. Consequently, this value is not comparable to its rarefied bootstrap distribution which is unbiased. Three estimations of probabilities, corresponding to the two-tailed hypothesis and the upper and lower one-tailed hypothesis, are computed. These probabilities correspond to the number of bootstrap replicates exceeding (upper one-tailed test), preceding (lower one-tailed test) the observed value divided by the number of bootstrap resamplings. For two-tailed probabilities, the bootstrap distribution is centered and the absolute values are used.

MDA runs without any function of optional MATLAB toolboxes. Only two freely available m-files of R. Bunschoten (distance.m) and R. Strauss (trilow.m) have been incorporated in the main function of the program (downloading them is not required). The association of both m-file permits a fast calculation of disparity metrics based on pairwise distances, especially for large sample size, comparing to code previously used.

3. Input files

Input files vary according to the analysis performed and user choices. For the MDA main function, data should be at least bidimensional and must be tab-delimited without headings (Table 2). One-dimensional or multiple one-dimensional data (each variable is treated separately and a maximum of 10 variables is analyzed) can be explored with MDA_1D function incorporating SGA (mean, minimum, maximum, variance and range of each variable) and BTailTest (on the variance of each variable) routines. MDA can be used with default filenames (contained in the MDA_prefs function). Where there are no default files, an input dialog box is opened. The number of files required ranges from two to four: (i) the first is for axis coordinates, (ii) the second for a matrix of presence/absence by stratigraphic level (or by geographical area, etc., depending on the subject-matter), (iii) an optional third input file corresponds to a further group file as the occurrence file allows multiple subdivision of the data file, (iv) the fourth and final file contains PC eigenvalues. This file is only required when a rescaling option is selected.

4. Output files and examples

Outputs consist in an ASCII file with its heading (which is also displayed in the MATLAB workspace)

Table 2
Example of format of input files

Observations	Data matrix.txt			Occurrence.txt					Group.txt			Eigenvalues.txt		
	Variables 1	*j*	*P*	levels 1	*j*			*L*	Group 1	*j*	*G*	Eigenv. 1	*j*	*P*
1	−0.851	2.228	−2.424	1	0	0	0	0	1	0	0	2.889	1.768	0.146
	−0.949	7.017	−6.161	1	0	0	0	0	1	0	0			
	2.515	8.239	−6.562	0	1	1	1	0	0	1	0			
i	−0.879	11.004	−2.699	0	1	1	1	0	0	1	0			
	−2.793	11.970	0.088	1	1	1	1	1	0	0	1			
N	−2.173	14.089	−7.431	0	0	0	1	1	0	0	1			

corresponding to the analysis summary and the user-choices, with the column headings and numerical results (Table 3). The first columns of numerical results correspond to sample size, rarefied size, interval and group (which is one for SGA). For the bootstrap tail test, the first columns correspond to the analyzed group and the observed value of the chosen disparity estimates. The Save option opens an output dialog box allowing the user to select a directory and filename. If the filename is already used, the results are appended to the prior results. At the end, the option: "return to user main page to run more analyses" or "exit" is displayed.

The program test and examples of output figures presented here are based on a stochastic simulation of the evolution of a monophyletic clade. The m-file is provided in the help folder of MDA (randclade.m) and is based in part on the Basic code of McShea (1994) for detecting major evolutionary trends. In our case, to simulate morphospace patterns, the code has been extended to two dimensions without boundaries and with the same probabilities for direction of change when anagenesis or cladogenesis occurs (unbiased model). The magnitude of change is equal in both dimensions and the probability of stasis is equal to the probability of change by anagenesis. At the end of the simulation, a subroutine is run in order to extract two monophyletic subclades in addition to the main clade. Simulation is performed with initial inputs based on empirical results obtained from a Cenozoic radiation of mammals (Alroy, 2000): the origination rate is 0.296 Lmyr and the extinction rate is 0.275 Lmyr (in line with the mean rate over 70 Ma from Alroy's data); the time span is 70 myrs. Arbitrarily, a step is understood as representing 0.2 myr and so simulation is performed in 350 steps. Steps are grouped into 14 bins, which represent durations of 5 myr. Then, a second subroutine is run to artificially alter the data and so simulate an incomplete, time-averaged fossil record. For each step, the species obtained are sampled randomly with a probability of preservation of 0.35 per myr (from the result for Cenozoic mammals reported by Foote and Raup, 1996). Preserved occurrences of species are regrouped within bins. The average morphology of a species encountered in all its sampled occurrences is used throughout its range. This latter

Table 3
Example of output files

Analysis summary

(A)
Analysis performs: SGA
Number of PCs retained: 2
Rescaling PCs on eigenvalues: *n*
Rarefaction: *y*: rarefaction size used: 4
Number of bootstrap resampling: 200
Upper-lower values: ±1stdev

Sample size	Rarefaction size	Interval	Group	Sum of ranges	Stdev SR	Upper value SR	Lower value SR
2	2	1	1	1.48	1.54	3.03	−0.06
4	4	2	1	9.35	2.66	12.01	6.68
13	4	3	1	11.83	4.80	16.64	7.03
10	4	4	1	15.53	5.54	21.08	9.99
13	4	5	1	22.06	7.41	29.47	14.65
14	4	6	1	22.18	8.91	31.09	13.27

(B)
Analysis performs: BTailTest
Number of PCs retained: 2
Rescaling PCs on eigenvalues: *n*
Rarefaction: none
Number of bootstrap resampling: 10,000
Upper-lower values: none
Disparity metric used: partial disparity

Group	Observed	*p* two-tail	*p* upper one-tail	*p* lower one-tail
1	327.09	0.9295	0.4649	0.5351
2	288.09	0.0178	0.9911	0.0089
3	199.28	0.000000	1	0.000000

subroutine is close to a classical method of sampling in disparity studies where one species per genus is sampled and used throughout the stratigraphical range of the genus (e.g., Eble, 2000a). So, from this simulation three data files (morphology, occurrence and group) are obtained and used as MDA input files.

4.1. Temporal patterns of disparity

In this case, the SGA option is used with (i) no rescaling of axes (since they represent crude characters), (ii) no CI, (iii) 200 bootstrap resampling iterations and (iv) a rarefaction size equal to the second minimum sample size observed ($n = 4$, $t = 65\,\mathrm{Ma}$). Two main figures are displayed. The first corresponds to the error bar plot of disparity estimates for 14 temporal bins (Fig. 1A). The 10 disparity descriptors display similar patterns of increase through time. However, range-based descriptors display a more monotonic pattern than the variance-based ones. The high error bar on the range-based descriptors is due to the small rarefaction size, which is a common problem when analyzing evolutionary radiations. The variance-based pattern exhibits more marked increases and plateau phases (near constant disparity). This pattern may be viewed as the reflection of modification of the morphospace structure with a discontinuous occupation of morphospace. Although descriptors tend to be correlated with each other, their responses to modifications of the morphospace, such as selective extinction (Foote, 1991), the splitting of morphospace occupation or bifurcation (Ciampaglio et al., 2001), can be very different and so a combination of different descriptors provides a clearer picture (Ciampaglio et al., 2001). So, discordances between disparity metrics can be observed, for example, after 25 Ma between the sum of variances, the root product of variances or the median pairwise distance. The second figure displayed shows the rarefied estimate of the location parameters on each axis of the morphospace (Fig. 1B). Estimates on character 1 show a first period where two estimates display similar behavior, and then, estimates display a disjunction in their behavior (a stasis for the Max and a decrease trend for the Min). This point of disjunction corresponds to the time of first net increase observed in some disparity estimates.

All of the patterns observed above relate to the diversification of the entire clade. But, the evolutionary history of a clade is a composite message of the history of subclades, which may exhibit, for example, relayed diversifications, competitive interactions, original morphological innovation, and different origination or extinction rates (see Jablonski, 2000, for a summary), thereby inducing specific evolutionary dynamics for each subclade and discordances with the pattern of the clade. Contrasting clade and subclade patterns account for the composite overall history. Disparities specific to the subclades can be viewed in terms of "internal" disparity (disparity specific to the subclade) and partial disparity (i.e., part of the overall disparity due to the subclade).

4.2. Temporal patterns of disparity of the two subclades

Internal disparity is analyzed in order to reveal patterns of diversification of selected subclades, as opposed to the entire clade. In this case, the MGA option is used with the same parameters of analysis as for SGA. Both main figures (obtained by SGA) are replicated for the number of groups in the input file (here two). Only one disparity estimate is presented here (the mean pairwise distance, Fig. 2A). The two subclades exhibit very different patterns from one another and from the clade (cf. Fig. 1A). For example, these patterns could be compared with patterns of diversity (Fig. 2B). Whereas the whole data set is random and that consequently no evolutionary or ecological forcing mechanism is operating, the two subclades mimic a replacement pattern (e.g., competition between the two subclades). Moreover, periods of diversification of each of the two subclades seem to support fluctuations in the clade overall.

4.3. Temporal patterns of partial disparity

A partial disparity analysis is performed in order to extract that part of disparity that can be attributed to subclades. In this case, the question is: which part of the clade's diversification is supported by which subclades? In this case, the PDA option is used with same initial parameters as in the previous two analyses but a CI of 95% is chosen, represented by an error bar. The same subclades analyzed previously are used. The figures (Fig. 3) plot temporal series of the partial disparity metric for each group. The first and last periods of the overall disparity seem to be supported by one of the two subclades (Fig. 3). The plateau phase does not seem to be explained by these two subclades. So, for their periods of domination, the two subclades seem to account for the majority of the pattern of overall disparity, and this despite their low diversity. So, it can be expected that species of these two subclades are located essentially in peripheral regions of the morphospace. In this case, their removal should induce a major decrease in the estimation of disparity.

4.4. Impact of the subclades on overall disparity

A bootstrap tail test is performed in order to estimate the impact of removing one subclade (the subclade 1 then the subclade 2) on overall disparity: the BTailTest option is used with 10,000 bootstrap replications on the partial disparity estimator. The groups are modified to represent species less the subclade species. The overall

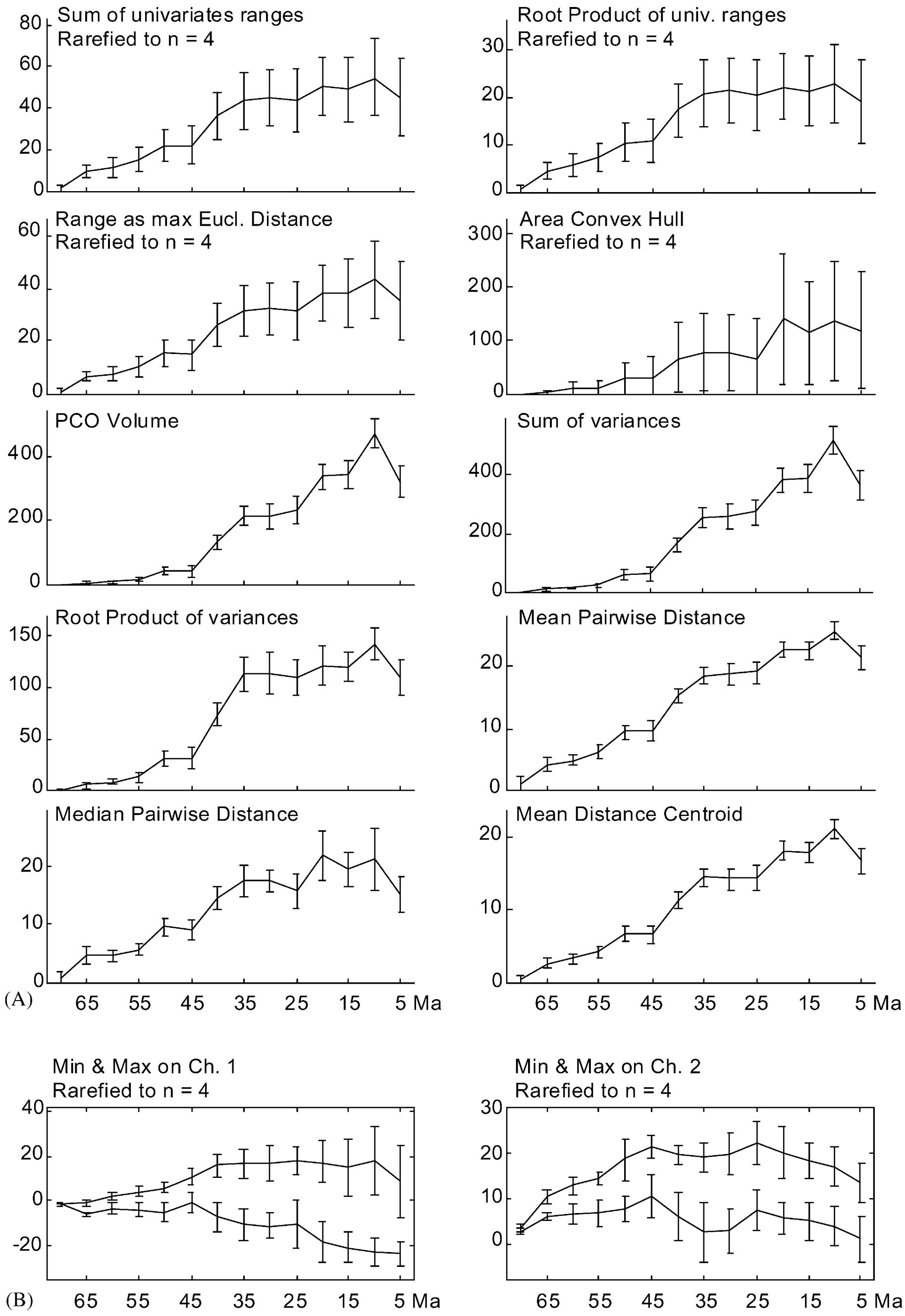

Fig. 1. Outputs from SGA corresponding to 10 estimates of disparity over time and two location estimates. (A) Disparity estimates. (B) Location estimates. Error bar represents ± one standard deviation calculated from 200 bootstrap replicates (rarefied to $n = 4$ or not).

morphospace is analyzed. The observed partial disparity of each group (the whole data set minus subclade 1 or 2) is compared with the bootstrap distribution (10,000 replicates) of the disparity of the entire clade using the partial disparity metric. Estimates of probabilities are displayed in the workspace if the "save numerical results" option is selected (Table 3B). Because disparity can either increase (omission of more centralized taxa in the morphospace than the larger group) or decrease (omission of morphologically peripheral group) when a

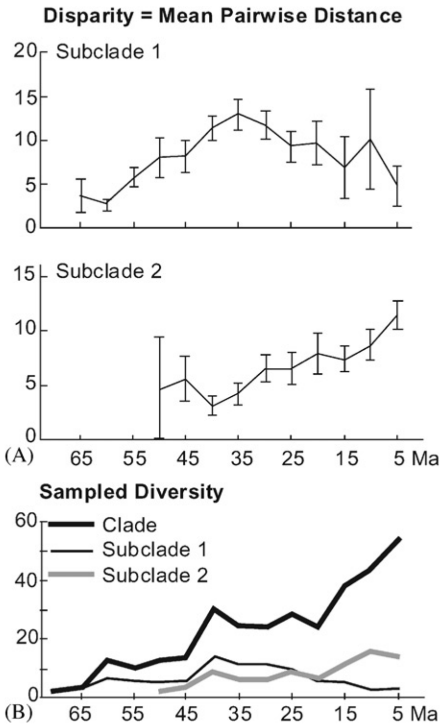

Fig. 2. Examples of outputs from MGA. (A) Disparity estimate is mean pairwise distance. Error bar represents ± one standard deviation calculated from 200 bootstrap replicates. (B) Diversity of three monophyletic clades and subclades (diversity equal to number of species sampled in bin).

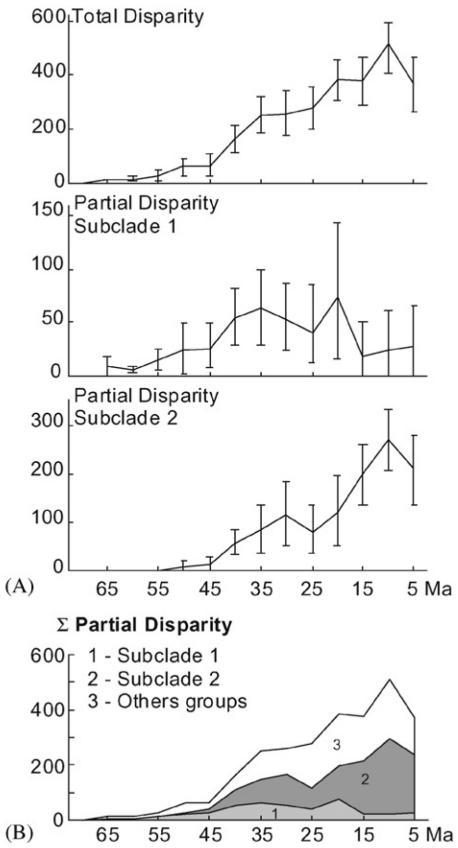

Fig. 3. Outputs from PDA corresponding to partial disparity over time of three clades. (A) Error bar represents a 95% CI calculated from 500 bootstrap replicates. (B) Disparity history of clade divided up into additive contributions made by two subclades selected and all others combined. This figure is not displayed initially but can be easily constructed from output data file.

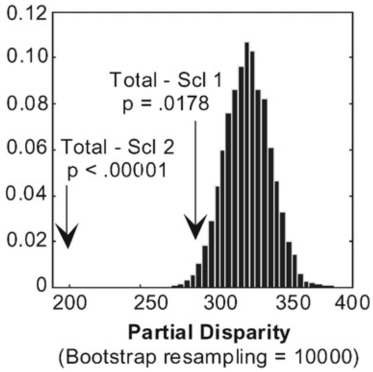

Fig. 4. Outputs from BTailTest. Bootstrap distribution of overall disparity (time averaging) and probabilities yielded by removal of one subclade and observed value of partial disparity. Test is based on 10,000 bootstrap replicates. Because disparity can either increase or decrease when a group is omitted, probabilities correspond to two-tailed tests.

group is omitted, all tests are two-tailed (Foote, 1993). As suggested by this author, tests in partial disparity only evaluate the probability that the observed effect of a group would disappear if the study is repeated. The results (Fig. 4) show that when the subclade 2 is omitted a significant decrease of disparity is observed. So, subclade 2 supports a major part of the overall disparity. This decrease in disparity shows that this subclade occupies an original peripheral morphological region within the morphospace. Similar results are obtained for subclade 1 but with lesser intensity. This can be explained by the fact that the period of diversification of this subclade is mainly at the beginning of the radiation when disparity is low. The importance of this clade in the early stages is dwarfed by the significance of the last period in the overall disparity.

5. Conclusion

The great development of morphospace-disparity analysis in paleobiological studies, and the challenge of extending them to other disciplines, require the development and use of various complementary metrics. It would allow to apprehend state-space patterns in terms of features such as dispersion, aggregation and location in order to extract complementary quantitative information about the way state-space is structured. The patterns obtained can then be contrasted with other biotic (e.g., extinction, origination rates) or abiotic (e.g., environmental proxy) information. State-space is sample dependent and so is biased by the analytical procedure. The MDA program offers various alternative tools for examining patterns of a state-space and procedures to correct for sample dependence. The versatility of MDA means that data can be explored quickly and the most interesting patterns extracted with minimum effort. The program can be easily applied to a wide range of problems involving state-space. In addition, this code can be modified easily and quickly for extension to other metrics.

Acknowledgements

I am grateful to D. McShea for providing me with the initial code of his 1994 paper and to R. Bunschoten and R. Strauss for providing m-files freely. I thank S. Montuire, P. Neige and T. Saucède for their comments on the initial version of the manuscript and my two reviewers J. Crampton and R. Lupia for their valuable suggestions for improving the manuscript and the program. This work was supported by a doctoral grant from the "Ministère de la Recherche" and is a contribution to the team C "Macroévolution et dynamique de la biodiversité" of UMR CNRS 5561-Biogéosciences.

References

Alroy, J., 2000. New methods for quantifying macroevolutionary patterns and processes. Paleobiology 26, 707–733.

Ciampaglio, C.N., Kemp, M., McShea, D.W., 2001. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. Paleobiology 27, 695–715.

de Vargas, C., Renaud, S., Hilbrecht, H., Pawlowski, J., 2001. Pleistocene adaptive radiation in *Globorotalia truncatuloides*: genetic, morphologic, and environmental evidence. Paleobiology 27, 104–125.

Dommergues, J.-L., Laurin, B., Meister, C., 1996. Evolution of ammonoid morphospace during the early Jurassic radiation. Paleobiology 22, 219–240.

Eble, G.J., 2000a. Contrasting evolutionary flexibility in sister groups: disparity and diversity in mesozoic atelostomate echinoids. Paleobiology 26, 56–79.

Eble, G.J., 2000b. Theoretical morphology: state of the art. Paleobiology 26, 520–528.

Eble, G.J., 2002a. Developmental morphospaces and evolution. In: Crutchfield, J.P., Schuster, P. (Eds.), Evolutionary Dynamics. Oxford University Press, Oxford, pp. 35–65.

Eble, G.J., 2002b. Multivariate approaches to development and evolution. In: McNamara, K., Minugh-Purvis, N. (Eds.), Human Evolution through Developmental Change. Johns Hopkins University Press, Baltimore, MD, pp. 51–78.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York, pp. 436.

Foote, M., 1991. Morphological and taxonomic diversity in a clade's history: the blastoid record and stochastic simulations. Contributions from the Museum of Paleontology, Vol. 28, University of Michigan, pp. 101–140.

Foote, M., 1992. Rarefaction analysis of morphological and taxonomic diversity. Paleobiology 18, 1–16.

Foote, M., 1993. Contributions of individual taxa to overall morphological disparity. Paleobiology 19, 403–419.

Foote, M., 1994. Morphological disparity in Ordovician–Devonian crinoids and the early saturation of morphological space. Paleobiology 20, 320–344.

Foote, M., 1995. Morphological diversification of Paleozoic crinoids. Paleobiology 21, 273–299.

Foote, M., 1997. The evolution of morphological diversity. Annual Review of Ecology and Systematics 28, 129–152.

Foote, M., 1999. Morphological diversity in the evolutionary radiation of Paleozoic and post-Paleozoic crinoids. Paleobiology 25, 1–115.

Foote, M., Raup, D., 1996. Fossil preservation and the stratigraphic ranges of taxa. Paleobiology 22, 121–140.

Hickman, C.S., 1993. Theoretical design space: a new program for the analysis of structural diversity. Neues Jahrbuch für Geologie und Paläontologie Abhandlungen 190, 183–190.

Jablonski, D., 2000. Micro-and macroevolution: scale and hierarchy in evolutionary biology and paleobiology. Paleobiology 264, 15–52.

Jackson, S.T., Overpeck, J.T., 2000. Responses of plant populations and communities to environmental changes of the late quaternary. Paleobiology 26, 194–220.

Jernvall, J., Hunter, J.P., Fortelius, M., 1996. Molar tooth diversity, disparity, and ecology in Cenozoic ungulate radiations. Science 274, 1489–1492.

Jernvall, J., Wright, P.C., 1998. Diversity components of impending primate extinctions. Proceedings of the National Academy of Sciences USA 95, 11279–11283.

Lupia, R., 1999. Discordant morphological disparity and taxonomic diversity during the cretaceous angiosperm radiation: North American pollen record. Paleobiology 25, 1–28.

McGhee, G.R., 1999. Theoretical Morphology. The Concept and its Applications. Columbia University Press, New York, pp. 316.

McShea, D.W., 1994. Mechanisms of large-scale evolutionary trends. Evolution 48, 1747–1763.

Norris, R.D., 2000. Pelagic species diversity, biogeography, and evolution. Paleobiology 26, 236–258.

Renaud, S., Michaux, J., Jaeger, J.-J., Auffray, J.-C., 1996. Fourier analysis applied to *Stephanomys* (rodentia, muridae) molars: nonprogressive evolutionary pattern in a gradual lineage. Paleobiology 22, 255–265.

Renaud, S., Benammi, M., Jaegger, J.-J., 1999. Morphological evolution of the murine rodent *Paraethomys* in response to climatic variations (mio-pliocene of North Africa). Paleobiology 25, 369–382.

Roy, K., Jablonski, D., Martien, K.K., 2000. Invariant size-frequency distributions along a latitudinal gradient in marine bivalves. Proceedings of the National Academy of Sciences USA 97, 13150–13155.

Smith, L.H., Bunje, P.M., 1999. Morphologic diversity of inarticulate brachiopods through the phanerozoic. Paleobiology 25, 396–408.

Smith, L.H., Lieberman, B.S., 1999. Disparity and constraint in olenelloid trilobites and the Cambrian radiation. Paleobiology 25, 459–470.

Wagner, P.J., 1996. Contrasting the underlying patterns of active trends in morphologic evolution. Evolution 50, 990–1007.

Wagner, P.J., 1997. Patterns of morphologic diversification among the Rostroconchia. Paleobiology 23, 115–150.

Wills, M.A., Briggs, D.E.G., Fortey, R.A., 1994. Disparity as an evolutionary index. A comparison of Cambrian and Recent arthropods. Paleobiology 20, 93–110.